#  STATISTIQUES DESCRIPTIVES

1. Caractéristiques d’une série statistique
2. Série statistique

Voici les séries de notes obtenues par 3 élèves :

*Jérôme :* 4 ; 6 ; 18 ; 7 ; 17 ; 12 ; 12 ; 18

*Bertrand :* 13 ; 13 ; 12 ; 10 ; 12 ; 3 ; 14 ; 12 ; 14 ; 15

*Julie :* 15 ; 9 ; 14 ; 13 ; 10 ; 12 ; 12 ; 11 ; 10

1. Moyenne

Méthode 1 : Calculer une moyenne

Calculer la moyenne pour chaque série de notes de Jérôme, de Bertrand et de Julie.

La moyenne est une caractéristique de position.

1. Médiane

Méthode 2  : Calculer une médiane

Calculer la médiane pour chaque série de notes de Jérôme, de Bertrand et de Julie.

Résumé : Pour déterminer les notes médianes, il faut ordonner les séries. La médiane partage l’effectif en deux.

Définition :

La **médiane** *m* est une valeur telle que la moitié au moins de l’effectif ait des valeurs inférieures ou égales à *m*, l’autre moitié des valeurs supérieures ou égales à *m*.

La médiane est une caractéristique de position.

1. Étendue

Définition : L’**étendue** d’une série statistique est la différence entre la plus grande valeur et la plus petite valeur de la série.

Méthode 3 : Calculer une étendue

Calculer l’étendue pour chaque série de notes de Jérôme, de Bertrand et de Julie.

(cas particulier de la série de Bertrand : élagage possible)

L’étendue est une caractéristique de dispersion.

1. Quartiles, écart interquartile

Définitions :

Le **premier quartile** est la plus petite valeur de la série telle qu'au moins 25 % des autres valeurs de la série sont inférieures ou égales à cette valeur.

Le **troisième quartile** est la plus petite valeur de la série telle qu'au moins 75 % des autres valeurs de la série sont inférieures ou égales à cette valeur.

Définition : L'**écart interquartile** d'une série statistique de premier quartile *Q1* et de troisième quartile *Q3* est égal à la différence *Q3* - *Q1*.

Remarque :

L'écart interquartile d'une série mesure la dispersion autour de la médiane. Il contient au moins 50% des valeurs de la série.

L'écart interquartile n'est pas influencé par les valeurs extrêmes de la série.

Méthode 4 : Calculer les quartiles

Calculer les quartiles pour chaque série de notes de Jérôme, de Bertrand et de Julie.

Résumé :

Pour déterminer les quartiles, il faut ordonner les séries.

Le premier quartile est la donnée de la série se trouvant au quart de l’effectif.

Le troisième quartile est la donnée de la série se trouvant au trois-quarts de l’effectif.

Les quartiles sont des caractéristiques de position.

L’écart interquartile est une caractéristique de dispersion.

1. Interprétations

*M(Jérôme)* = 11,8 *m(Jérôme)* = 12 *E(Jérôme)* = 14 *Q1(Jérôme)*  = 6 *Q3(Jérôme)* = 17

 *EQ(Jérôme)* = 11

*M(Bertrand)* = 11,8 *m(Bertrand)* = 12,5 *E(Bertrand)* = 5 *Q1(Bertrand)*  = 12 *Q3(Bertrand)* = 14

 *EQ(Bertrand)* = 2

*M(Julie)* ≈ 11,8 *m(Julie)* = 12 *E(Julie)* = 6 *Q1(Julie)*  = 10 *Q3(Julie)* = 13

 *EQ(Julie)* = 3

Les moyennes sont environ égales et pourtant les notes ne se répartissent pas de la même manière autour de cette caractéristique de position. Les étendues sont très différentes.

Dire que Jérôme à une médiane égale à 12 signifie que Jérôme a obtenu autant de notes au-dessus de 12 que de notes en-dessous de 12.

Dire que le premier quartile de Bertrand est égal à 12 signifie qu’au moins un quart des notes de Bertrand sont inférieures à 12.

Dire que le troisième quartile de Julie est égal à 13 signifie qu’au moins trois quarts des notes de Julie sont inférieurs à 13.

L’écart interquartile de Jérôme est égal à 11 signifie qu’au moins 50% des notes de Jérôme sont comprises entre 6 et 17 (les quartiles).

1. Cas de pondération d’une série statistique
2. Série statistique

Tailles des élèves de 2nde5 en cm :

*174 – 160 – 161 – 166 – 177 – 172 – 157 – 175 – 162 – 169 – 160 – 165 – 170 – 152 – 168 – 156 – 163 – 167 – 169 – 158 – 164 – 151 – 162 – 166 – 156 – 165 – 179*

1. Regroupement par classe

Regrouper cette série de tailles par classes de longueur 5 cm et calculer les fréquences arrondies au centième :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Tailles | 150 $\leq $ t < 155 | 155 $\leq $ t < 160 | 160 $\leq $ t < 165 | 165 $\leq $ t < 170 | 170 $\leq $ t < 175 | 175 $\leq $t < 180 |
| Effectifs |  |  |  |  |  |  |
| Fréquences |  |  |  |  |  |  |

*L’effectif total est 27.*

Méthode 5 : comment réaliser un Histogrammes

Faire un histogramme en suivant la répartition proposée dans le tableau

Remarque :

Dans un histogramme, l’aire des rectangles est proportionnelle à l’effectif (ou à la fréquence).

Pour illustrer ça, on va dessiner à côté de l’histogramme précédent un autre représentant les mêmes données mais en fusionnant les deux premières classes.

1. Moyenne pondérée

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Classes centrées *xi* |  |  |  |  |  |  |
| Effectifs *ni* |  |  |  |  |  |  |

Il s’agit d’un calcul de moyenne pondéré car des effectifs différents *ni* sont associés à chaque valeur *xi*.

Définition :

La **moyenne** d’une série statistique dont les valeurs sont *x1*, *x2* , …, *xk* et les effectifs correspondants *n1*, *n2*, …, *nk* est notée $\overbar{x}$ et est égale à $\overbar{x}=$ $\frac{n\_{1}x\_{1}+…+n\_{k}x\_{k}}{n\_{1}+…+n\_{k}}$ ..

Ainsi dans l’exemple :

$\overbar{x}$ *= (2 x 152 + 4 x 157 + 7 x 162 + 8 x 167 + 3 x 172 + 3 x 177) : 27 = 4449 : 27*

 ≈ *164,8 cm*

Remarque : Calcul de la moyenne exacte

*(174 + 160 + 161 + 166 + 177 + 172 + 157+ 175 + 162 + 169 + 160 + 165 + 170 + 152 + 168 + 156 + 163 + 167+ 169 + 158 + 164 + 151 + 162 + 166+ 156 + 165 + 179) : 27*

*= 4444 : 27*  ≈ *164,6 cm*

La méthode de calcul de moyenne en centrant les classes est très fiable *(ici : 2 mm d’erreur)*

1. Linéarité de la moyenne

Propriété : Si une série de valeurs $x\_{i}$ a pour moyenne $\overbar{x}$, alors la série de valeurs $ax\_{i}+b$, avec *a* et *b* réels, a pour moyenne $a\overbar{x}+b$.

Exemple :

 **Moyennes**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| $$x\_{i}$$ | 4 | 7 | **–**2 | **(4 + 7 – 2) : 3 = 3** |
| $$2x\_{i}-5$$ | 3 | 9 | **–**9 | **(3 + 9 – 9) : 3 = 1** |

On a $\overbar{x}=3$, et donc en appliquant la propriété, la moyenne de la série $2x\_{i}-5$ est égale à : $2\overbar{x}-5=2×3-5=1$. On retrouve bien le résultat calculé directement dans le tableau.

1. Variance, écart-type

Définitions : - La **variance** *V* d'une série statistique de moyenne $\overbar{x}$ dont les valeurs du caractère sont *x1*, *x2*, *x3*, …, *xk* et les effectifs correspondants sont *n1*, *n2*, *n3*, …, *nk* est égale à : $V=$ $\frac{n\_{1}×\left(x\_{1}-\overbar{x}\right)^{2}+n\_{2}×\left(x\_{2}-\overbar{x}\right)^{2}+…+n\_{k}×\left(x\_{k}-\overbar{x}\right)^{2}}{n\_{1}+n\_{2}+…+n\_{k}}$ .

- L'**écart-type** σ d'une série statistique de variance *V* est égal à : $σ=\sqrt{V}$.

Ainsi en reprenant l’exemple des tailles de la 2nde5, la variance est égale à :



$$σ≈\sqrt{46,914}≈6,85$$

L'écart-type possède la même unité que les valeurs de la série.

Ainsi pour la série étudiée, l'écart-type est environ égal à 6,85 cm.

Remarque :

L'écart-type exprime la dispersion des valeurs d'une série statistique autour de sa moyenne. Les valeurs extrêmes influencent l'écart-type.

Bonus : petit point sur la médiane et les quartiles

#### Médiane

**Définition**

Soit une série statistique d'effectif total *n*, rangée par ordre croissant.

On appelle médiane la valeur "du milieu". On dit qu'elle partage la série en deux moitiés : il y a autant de valeurs en dessous qu'au dessus.

Pour déterminer son rang, il y a 2 cas :

|  |  |  |
| --- | --- | --- |
| * si *n* est impair : la médiane est la valeur de rang
 | 1843444662 | méd = 44 |
| * si *n* est pair : nous prendrons la demi-somme des deux valeurs dont les rangs entourent le nombre
 | 184344466295 | méd = 45 |

**Remarque :**

Si les données ont été regroupées en classes, on ne peut déterminer la valeur exacte de la médiane. En revanche, on appellera classe médiane, la classe qui la contient (et permet donc d'en donner un encadrement).

**Exemples**

Données discrètes "en vrac"

 10, 7, 12, 18, 16, 15, 5, 11, 11, 20, 15, 11, 18, 14

Ordonnons la série par ordre croissant : 5, 7, 10, 11, 11, 11, 12, 14, 15, 15, 16, 18, 18, 20

Il y a 14 termes or = 7,5.

La médiane est donc la demi somme des 7ème et 8ème termes : méd = = 13

Avec un tableau d'effectifs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| valeurs | 1 | 2 | 3 | 4 | 5 | 6 |
| effectifs | 6 | 11 | 25 | 19 | 15 | 5Attention, il faut bien interpréter cette dernière ligne : Les données qui valent 3 ont un rang compris entre 18 et 42 inclus |
| effectifs cumulés | 6 | 17 | 42 | 61 | 76 | 81 |

L'effectif total est de 81 or = 41. La médiane est donc le 41ème terme : méd = 3

Avec des données réparties par classes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| classe | [0 ; 2[ | [2 ; 4[ | [4 ; 6[ | [6 ; 8] |
| fréquence | 10% | 38% | 45% | 7% |
| fréquence cumulée | 10% | 48% | 93% | 100% |

48% des valeurs sont strictement inférieures à 4

Et 93% des valeurs sont strictement inférieures à 6

La classe médiane est donc la classe [4 ; 6[

On peut donc en déduire l'encadrement suivant 4 méd < 6

#### III. Quartiles

**Définitions :**

Le premier quartile $Q\_{1}$est la plus petite des valeurs de la série telle qu’au moins 25% de la population ait sa valeur inférieur ou égale à $Q\_{1}$.

Le troisième quartile $Q\_{3}$est la plus petite des valeurs de la série telle qu’au moins 75% de la population ait sa valeur inférieure ou égale à $Q\_{3}$.

**Méthode pour trouver** $Q\_{1}$ **et** $Q\_{3}$**.**

Pour une population d’effectif n, $\frac{n}{4}$ (ou $\frac{3n}{4}$) si il est un entiers nous donne le rang de $Q\_{1}$ (ou $Q\_{3}$), dans le cas contraire on prendra la valeur de l’élément du rang immédiatement supérieur.

**Exemple**

Avec un tableau d'effectifs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **valeurs** | 1 | 2 | 3 | 4 | 5 | 6 |
| **effectifs** | 6 | 11 | 25 | 19 | 15 | 5 |
| **effectifs cumulés** | 6 | 17 | 42 | 61 | 76 | 81 |

Ici n = 81 Donc $\frac{n}{4}=20,25$ il va donc falloir que je prenne la valeur de l’individu de rang 21, donc $Q\_{1}=3$

 $\frac{3n}{4}=60,75$ L’individu de rang 61 a pour rang 5 donc $Q\_{3}=4$

Exercice 1



Exercice 2



Exercice 3



Exercice 4



Exercice 5



Exercice 6





Exercice 7





Exercice 8



Exercice 9



Utiliser Excel pour faire des statistiques :

**Pour faire une légende variable :**

les parties fixes s’écrivent entre guillemets, la concaténation se fait avec l’esperluète (le symbole &).

**Pour faire dénombrer les cellules vérifiant un ou plusieurs critères**

Il faudra taper =nb.si.ens(cellules à explorer ; condition 1 à écrire entre guillemet ; cellules à explorer ; condition 2 ; ….)

Pour indiquer les cellules à explorer on peut taper l’information à la mains de la manière suivante : référence cellule supérieure gauche de la zone à couvrir : référence cellule inférieure droite de la zone à couvrir) , par exemple : B8 :D27 si on veut indiquer toutes les valeurs dans la zone allant de la ligne 8 à la 27 et allant de la colonne B à la D.

Comme condition on indiquera par exemple « >=8 » si on a envie de sélectionner uniquement des cellules contenant des valeurs supérieures à 8.

Par exemple

**Pour créer des valeurs aléatoires entre deux bornes :**

=valeur minimale + ENT(ALEA()\*(valeur maximale – valeur minimale))+1

=ALEA.ENTRE.BORNES(valeur minimale;valeur maximale)

Recherche de la médiane

Si on a un paquet de valeurs non classées il faudra utiliser la fonction mediane( comme suit :

« =MEDIANE(valeurs dont on cherche la médiane) »

Sinon il va falloir bricoler en utilisant les effectifs cumulés croissants

**fonctions de base :**

pour arrondir à l’entier supérieur (pratique pour les quartiles) : arrondi.sup(valeur, 0)

pour faire la somme de plusieurs valeurs : somme(valeurs)

**Ressources :**

<https://www.cours-gratuit.com/tutoriel-excel/tutoriel-excel-comment-compter-le-nombre-de-cellule-comprise-entre-deux-valeurs>

pour faire un polygone d’effectif ou de fréquence cumulées croissantes : <https://www.youtube.com/watch?v=VaRUSQmwN8I>